Noninformative prior in the quantum statistical model of pure states and generalized entropy (New developments of generalized entropies by functional analysis)

田中, 冬彦

数理解析研究所講究録
第 1852 巻 2013 年 192-204

Noninformative prior in the quantum statistical model of pure states and generalized entropy

東京大学・情報理工学系研究科 田中冬彦 (Fuyuhiko Tanaka)
The University of Tokyo

Abstract

In the quantum state tomography, which is the task of estimating the unknown quantum state on a Hilbert space from measurement data, the standard Bayesian technique is often used. As is well known in Bayesian statistics, it is a nontrivial problem to define a noninformative prior in general statistical models. Although some definitions for statistical models of mixed states are given by straightforward extension of classical results, there is no counterpart for those of pure states. Recently the author proposed a conceptual extension of the famous reference prior by Bernardo and defined noninformative priors for statistical models of pure states. While the formal extension of the reference prior corresponds to the prior maximizing the von Neumann entropy, our proposed prior is shown to maximize the limit of the quantum Renyi entropy. We show some relations between two principles.

1 Introduction

近年, 量子力学の原理の工学的応用を目指した実験や, 実験技術そのもの進歩により実験データの効率的・高精度の解析のため現代的な統計的手法が用いられつつある [25]. 特に, 量子トモグラフィでは, 科学的な法則性を明らかにするという目的ではなく, 工学的応用に向けた統計的手法の適用に興味がシフトしている. 例えばベイズ統計 [9, 29, 8] やモデル選択 [32] などが挙げられる.

通常, ベイズ統計学では確率分布のパラメータ族を考え, パラメータにも確率分布 (事前分布) を設定する (ベイズ統計一般については例えば Robert [27] を参照.). データ数がさほど多くなくても安定した推定ができるなどのメリットがある. しかし一方で, ベイズ統計はパラメータに関して全く情報がないような場合に事前分布をどうとるべきかという問題がある. (e.g., Kass and Wasserman [22].) ラフに述べるとデータ数が多くなれば, 推定における事前分布の影響は小さくなるが, データ数が少ない時には事前分布の取り方の影響を強く受けてしまう. そこで, パラメータに関する余計な情報 (もしくは解析する者の先入観) をできるだけ排した事前分布が望ましい. その意味で無情報事前分布 (Noninformative prior の訳語; vague prior, objective prior などもいう) を設定したいのだが, これは難しい問題である. パラメータの動く範囲で「一様分布」にする, あるいは対称性をもつモデル

であれば, 対称性を用いて事前分布を与えるというような素朴なやり方では問題が起きることはベイズ統計ではよく知られている.

連続パラメータをもつ多くの統計モデルでは正則条件を満たすためいわゆる Jeffreys prior [21] をはじめとして, 微分幾何学 (情報幾何学 [2]) の観点から事前分布を定義できる. これらはパラメータの変換に対して不変であるように定義される. しかし, パラメータが離散的な場合には幾何学的な観点からのアプローチはできなくなる. そのような場合, 理論的には Bernardo による reference prior [4] が古くから知られている. 適当な正則条件の下で, データ数が十分多いという近似の下では reference prior は Jeffreys prior に一致することも知られている [10]. Reference prior の定義の背後にあるのは情報理論における確率分布の符号長解釈である. 符号長解釈の下ではエントロピーは理想符号長の期待値, 相対エントロピーは理想符号長と推定した符号長との差の期待値という解釈ができる. このような解釈は情報理論以外にも, 例えば, 機械学習 [7] や最小記述長原理 [14] にも深く関わっている.

量子系の統計的推測にベイズ統計を用いる場合にも, 形式的なアナロジーから Jeffreys prior や reference prior を考えることは可能である. しかし, 古典ベイズ統計に比べて, 統計的推測への応用という観点での研究はまだほとんどないのが現状である. また, 量子統計モデルに対する無情報事前分布は統計的な応用とは無関係に古くは量子統計力学, 最近では量子情報理論や量子情報幾何の観点からも議論されている. (これらの関連研究は著者による文献 [30] を参照.) このように量子系での無情報事前分布は量子物理の実験だけでなく, 量子論の理論的な側面や量子情報幾何など様々な分野と密接に関連している.

以上のような状況を踏まえ, Tanaka [30] では量子系特有の統計モデルとして純粋状態モデルの無情報事前分布について考察している. 純粋状態モデルは古典論的な不確定性が全くない統計モデルのため, reference prior を形式的に拡張して使う積極的な理由はない. Tanaka は符号長解釈の形式的な拡張ではなく, むしろ reference prior の根本にあるゲーム理論的なアイディアに注目し, 純粋状態モデルの場合に全く異なる事前分布 (MDP prior) を定義した.

本稿では, 以上の結果を踏まえ, MDP prior が, Rényi entropy の極限を最大化するという方法で得られることを示す. また, reference prior の形式的な拡張は, Holevo 情報量の最大化になるが, 純粋状態モデルでは von Neumann entropy の最大化で得られる. つまり, 2 種類の無情報事前分布について, それぞれの背後にエントロピー的な量が付随していることが明らかになった. 本稿では, これらの比較に関する解析結果の一部を紹介しておく.

まず 2 節では reference prior の定義と符号長解釈について簡単に説明する. 3 節では形式的な拡張で量子系にも reference prior が定義されることを示す. 次に 4 節で純粋状態族における reference prior の概念的な問題を説明し, 著者により提案された MDP prior を紹介する. 5 節では一般化エントロピーの視点から両者をとらえなおす.

本研究集会のキーワードであるエントロピー概念が様々な分野に浸透していることと同様, 本稿における結果の背後にも, 情報理論やベイズ統計学, 量子論など幅広い分野の話題が関わっている. 基本的な事柄について最低限の説明を入れているものの, 詳しくはそれぞれの標準的な教科書を参照していただきたい.

2 Reference prior

本節では古典統計における reference prior について定義を確認しておく. ベイズ統計と無情報事前分布については例えば Robert [27] などを参照.

以降, x は観測されるデータとし, 未知パラメータ θ をもつ確率分布 $p(x|\theta)$ (もしくは p_θ などとも書く) から発生していると仮定する. これは $x \sim p(x|\theta)$ のように書く. パラメータ θ は k 次元ユークリッド空間の Borel 集合 $\Theta \subseteq \mathbf{R}^k$ を動くものとする. データの取りうる値は有限個 $\mathcal{X} = \{x_1, \ldots, x_k\}$ としておく. この時, $\forall\theta \in \Theta$ で $\sum_x p(x|\theta) = 1$, $p(x|\theta) \geq 0, \forall x \in \mathcal{X}$ が成立. 確率分布の族 $\mathcal{M} = \{p(x|\theta) : \theta \in \Theta \subseteq \mathbf{R}^k\}$ を統計モデルと呼ぶ. 本稿では統計モデルでは常に $\theta \neq \theta' \Rightarrow p_\theta \neq p_{\theta'}$ を仮定する.

ベイズ統計学の立場では, θ も確率変数と見るため $\pi(\theta)$ なる確率分布を仮定する. ($\int \pi(\theta)\mathrm{d}\theta = 1, \pi(\theta) \geq 0$.) (記述の簡単のため Lebesgue 測度 $\mathrm{d}\theta$ に関する絶対連続性を仮定しておく.) これを事前分布 (prior distribution; prior) などと呼ぶ. 事前分布の集合は $\mathcal{P}(\Theta)$ と表す. 理論的には積分が規格化できない improper prior がしばしば出てくるが, ここでは, 積分有限 (従って規格化できる) の事前分布 (proper prior) のみ考えることにする.

さて, 二つの確率変数 x, θ の相互情報量は

$$
\begin{aligned}
I_\pi := I(X:\theta) &:= \int \pi(\theta) \int p(x|\theta) \log \frac{p(x|\theta)}{p_\pi(x)} \mathrm{d}\theta \\
&= \int \pi(\theta) D(p_\theta || p_\pi) \mathrm{d}\theta \qquad (1) \\
&= \int p_\pi(x) D(\pi(\theta|x) || \pi(\theta)) \mathrm{d}x \qquad (2)
\end{aligned}
$$

のように定義される. (詳しくはたとえば, Cover and Thomas [11] を参照.) ここで $p_\pi(x) := \int p(x|\theta)\pi(\theta)\mathrm{d}\theta$ and $\pi(\theta|x) := \frac{p(x|\theta)\pi(\theta)}{p_\pi(x)}$ はそれぞれ, データ x の周辺分布とパラメータ θ に関する事後分布である.

Bernardo [4] は (2) の形に注目し, 統計モデルが与えられた下で相互情報量を事前分布 π の汎関数とみなして

$$\pi_{REF} := \arg\max I(\pi)$$

を無情報事前分布として用いることを提案した．これは reference prior などと呼ばれベイズ統計に基く様々な統計解析に応用されている [5].

相互情報量の解釈としては，データ x を得た時に知りえるパラメータ θ の情報量を考えて，その期待値をとっている．実際 x, θ が独立の場合，つまり $p(x|\theta)$ が θ に依存しない場合には相互情報量が 0 になる．一方で，量子系への拡張には (1) の形を用いる必要がある．こちらの書き方は符号長解釈を前面に押し出した形になる．そこで，以下，確率分布の符号長解釈について簡単に説明する．

データ x が確率分布もしくは相対頻度 $p_\theta(x)$ に従って独立に発生しているとみなせるとき，θ を仮に知っていれば，理想的な符号長は $-\log p_\theta(x)$ で与えられる．（実際には，log の底を 2 にとり，$L(x) = \lfloor -\log_2 p_\theta(x) \rfloor + 1$ であるが，ここでは概念的な考察のため簡略化している．）一方で θ を知らない場合には $q(x)$ なる分布を想定して 符号長として $-\log q(x)$ を各データ x に割り当てる．この場合，符号長の差 (冗長度) は

$$-\log q(x) - (-\log p_\theta(x))$$

で与えられる．データの分布 $p_\theta(x)$ に関する期待値をとると

$$\mathrm{E}[-\log q(x) - (-\log p_\theta(x))] = D(p_\theta||q)(\geq 0)$$

のように p_θ と q の相対エントロピー（損失関数）を得る．q としてはすべての θ に対して，できるだけ上の損失関数を小さくできるものが望ましい．今，θ が不明でも $\pi(\theta)$ がわかっている場合には，平均的に上の損失関数を最小化するような $q(x)$ を考えることができ，それは，$-\log p_\pi(x)$, $p_\pi(x) := \int p_\theta(x)\pi(\theta)\mathrm{d}\theta$ で与えられる．つまり，

$$D(p_\theta||p_\pi) = \min_{q(x)} \int D(p_\theta||q)\pi(\theta)\mathrm{d}\theta$$

が成立する [1].

情報理論では $-\log p_\pi(x)$ による符号を π に関するベイズ符号と呼ぶ．適当な正則条件の下で

$$\begin{aligned}\sup_\pi I(\pi) &= \sup_\pi \inf_q \int D(p_\theta||q)\pi(\theta)\mathrm{d}\theta \\ &= \inf_q \sup_\pi \int D(p_\theta||q)\pi(\theta)\mathrm{d}\theta\end{aligned}$$

が成立し sup, inf は max, min で置き換えてよいことが知られている．二番目の等号はもっと一般の問題でも成立しミニマックス定理と呼ばれる．特に reference prior π_{REF} に基いたベイズ符号はミニマックス符号にもなっていることがわかる [12, 15].

なお，$C = \max_\pi I(\pi)$ は通信路容量とよばれ情報理論では重要な量になっている．

3 量子系への形式的な拡張

良く知られているように, 量子情報理論では情報源は確率分布ではなく密度作用素になる. 量子系を表すヒルベルト空間$\mathcal{H}$について, 量子状態は条件$\mathrm{Tr}\rho = 1, \rho \geq 0$を満たす線形作用素$\rho$で記述される. (以降, しばしば適当な正規直交系を固定して行列とみなす. 本稿では可分ヒルベルト空間のみ扱い非有界作用素も出てこない. そのため行列のアナロジーで理解してもほとんど問題ない) このような線形作用素ρを密度作用素という. 密度作用素の全体を$\mathcal{S}(\mathcal{H})$で表す. 特にランク1の密度作用素, つまり$\rho = |\psi\rangle\langle\psi|$のようにベクトルの外積の形で書ける密度作用素を純粋状態と呼ぶ.

データの確率分布は, 量子状態ρを情報源とみなした時, ρに対する測定方法をひとつ定めることで決まる. (詳しくは量子情報の標準的な教科書である Nielsen and Chuang [23] などを参照.)

密度作用素ρは定義から$\rho = U\mathrm{diag}(\lambda_1, \lambda_2, \ldots)U^*$のように表示すれば$\sum_j \lambda_j = 1$, $\lambda_j \geq 0$のようにかけるから離散的な確率分布と同一視できる. したがって, 測定方法を定めずとも, 古典情報理論の話題はある程度, 数学的なアナロジーで拡張することもできる. たとえば, 量子相対エントロピーなどは, その操作論的意味付けが明確になる前に数学的には以下のように与えられていた.

$$D(\rho||\sigma) := \mathrm{Tr}\rho(\log\rho - \log\sigma).$$

これは$\rho = U\mathrm{diag}(p_1, p_2, \ldots)U^*, \sigma = U\mathrm{diag}(q_1, q_2, \ldots)U^*$のような場合には通常の相対エントロピーに帰着する. 他にも量子相対エントロピーは確率分布に対する相対エントロピーと同様の性質を幾つか持っていることが示されている. 例えば, $D(\rho_1||\rho_2) \geq 0$であって$D(\rho_1||\rho_2) = 0 \Leftrightarrow \rho_1 = \rho_2$もいえる (ただし, 証明は簡単ではない.) 量子相対エントロピーについて詳しくは, 例えば Ohya and Petz [24] を参照せよ.

さて, 密度作用素のパラメータ族 (量子統計モデルという) $\{\rho_\theta : \theta \in \Theta \subseteq \mathbf{R}^k\} \subset \mathcal{S}(\mathcal{H})$が与えられている時には, パラメータの事前分布$\pi$に関して

$$I_Q(\pi) := \int \pi(\theta) D(\rho_\theta||\rho_\pi)\mathrm{d}\theta$$

なる汎関数が定義できる. ただし, $\rho_\pi := \int \rho_\theta \pi(\theta)\mathrm{d}\theta$. そして, Bernardo による reference prior の形式的な拡張は$\pi_{REF} := \arg\max I_Q(\pi)$で与えられる. なお, 上の量は量子情報では一般に Holevo 情報量 [17] と呼ばれている.

純粋状態からなる量子統計モデルの場合にも形式的にこのような定義を用いることもできるが, 後で見るように解釈上の問題がある.

また, Bernardo の本来の着眼点は相互情報量の (2) の形であるが, 事後分布は測定方法をひとつ特定しないと定まらない. 量子統計モデルの reference prior に関

しても統計学的な意味は詳しく調べられていないため, これらの解析も今後の課題である.

4 純粋状態モデルの場合

量子系の統計的推測では, 量子系を記述する密度作用素を測定データから推定する問題を取り扱う. 適切な測定方法の下でデータ数 N が無限にあれば, いくらでも精度よく推定できる. しかし, 系を指定するパラメータの次元 k が大きい状況では, 実用上・効率上, 相対的にデータ数 N が少ない状況 $N \sim k$ で推定する必要がある. 従って, データ数は計算機の性能も考慮して十分限られていることを前提とした理論が必要になる.

本稿の以下の内容には不要だが, 量子統計の研究に関する文献は最近, かなり増えてきたため関連する文献として幾つか挙げておく. まず, 内容は古いものの Helstrom [16] や Holevo [18] が基本的な教科書である. ただし, Holevo は理論的な興味, 例えば推定精度の原理的な限界といった興味が中心である. 実験な応用に関する話題は Paris *et al.* [25] が詳しい. 統計研究者には, 量子光学でのトモグラフィに限定されるが, Artiles *et al.* [3] が数値結果も図示しておりわかりやすい. また, これらの問題は, 量子情報理論では量子推定と呼ばれているが, 数理統計的な話題とは大きく異なるものも多いように思われる. (このあたりの物理分野と統計分野での誤解については著者による解説 [28] に詳しくまとめている.)

本稿では純粋状態からなる量子系に注目する. これは古典的な不確定性が全くない状態に相当する. 応用としては, 有限次元パラメータをもつ波動関数のベイズ推定などが考えられるだろう. 波動関数もしくは純粋状態はランク 1 の密度作用素だから, これらについて考察することは, 一般にランク r の密度作用素という制限の下でベイズ推定を考える場合の足がかりにもなりえる. 量子光学での光の状態を記述する場合, 一般に密度作用素の成分が無限個になるが, 実際には, 有限次元の行列に落として推定している [3]. 従って, このような場合のベイズ推定を研究する上でも重要である.

我々の関心は, 純粋状態族 $\mathcal{M} := \{\rho_\theta := |\psi_\theta\rangle\langle\psi_\theta| : \ \theta \in \Theta \subseteq \mathbf{R}^k\}$ におけるパラメータ θ の無情報事前分布である. まず, 前節で導入した reference prior を考えてみよう. 純粋状態のエントロピーについて $S(\rho_\theta) = 0$ に注意すれば, Holevo 情報量は以下のように変形できる.

$$\begin{aligned}\int D(\rho_\theta || \rho_\pi)\pi(\theta)\mathrm{d}\theta &= -\int \mathrm{Tr}\rho_\theta \log \rho_\pi \pi(\theta)\mathrm{d}\theta - \int S(\rho_\theta)\pi(\theta)\mathrm{d}\theta \\ &= S(\rho_\pi).\end{aligned}$$

従って, reference prior はエントロピー最大化で得られることになる.

エントロピーを我々が系に対してもつ不確定性とみなすと, 情報が全くない場合は, 不確定性を最大にするのが自然に思える. 古典的な場合には, Jaynes がエネルギーのアンサンブル平均を固定した下でエントロピーを最大化する分布として Boltzman 分布が導出されることを示している [19, 20]. このような考え方は最大エントロピー原理, もしくは Jaynes 原理などと呼ばれ, 確率密度関数の推定などでも使われている. 従って, $S(\rho_\pi)$ を最大化する事前分布を無情報事前分布とするのは, Jaynes 原理を形式的に量子論的なケースに拡張していると考えることもできる.

しかし, 純粋状態の族に対して, 我々が系に対してもつ不確定性というのはユニタリ方向の不定性である. たとえば, 以下のように未知パラメータ $s, t \in \mathbf{R}$ を持つ純粋状態モデルを考えてみよう.

$$\rho_{s,t} := |\varphi(s,t)\rangle\langle\varphi(s,t)|,$$
$$|\varphi(s,t)\rangle := U(s,t)|\varphi_0\rangle,\ U(s,t) = e^{-i(sA+tB)},$$

ここで A, B は互いに非可換な observable である. 異なる $(s,t), (s',t')$ に対して純粋状態ベクトル $|\varphi(s,t)\rangle, |\varphi(s',t')\rangle$ はお互いに直交しない. 我々が s, t といったパラメータを知らないことによる不定性はエントロピーの導出で考えていた不定性とは違っていることに注意すべきである. エントロピーには様々な特徴づけがあるが, その一つは互いに識別可能な状態 $\{1, \ldots, k\}$ に対して, 組み合わせの数を数え上げる所から来ていた. 量子論であっても, 系のハミルトニアンに関するエネルギー固有状態達, $|E_1\rangle, |E_2\rangle, \ldots$ は直交していた (識別可能) ことを思い出しておく. $|\varphi(s,t)\rangle$ のように識別できない非直交な場合にそのまま適用してよいという根拠はない.

また, Bernardo の方法の根本に戻ると, そこには符号長解釈があった. しかし, 純粋状態モデルでは形式的に $-\log\rho$ という "符号長" 作用素を考えた場合, 符号長として 0 か ∞ を固有値にもつ作用素ができる. 特にエントロピーや相対エントロピーは

$$S(|\psi\rangle\langle\psi|) = 0,\ D(|\psi\rangle\langle\psi| \,||\, |\phi\rangle\langle\phi|) = \infty \quad \text{if } \psi \neq \phi$$

となってしまう. 後者が致命的で $|\langle\psi_1|\psi_2\rangle| = 0.01$, $|\langle\psi_1|\psi_3\rangle| = 0.99$ のような場合であっても $D(|\psi_1\rangle\langle\psi_1| \,||\, |\psi_2\rangle\langle\psi_2|) = D(|\psi_1\rangle\langle\psi_1| \,||\, |\psi_3\rangle\langle\psi_3|) = \infty$ となる. これはエントロピー的な概念が識別不可 (非直交) な二つの純粋状態の違いをうまく測れていないことに起因すると考えられる. こういった事情から, 著者は Bernardo の考え方における符号長解釈を捨てて, むしろ, ゲーム理論的な解釈に注目した. その結果, 純粋状態モデルに特有のゲームを設定することで事前分布を構成できることを示した [30]. もちろん, ゲームの設定次第で事前分布も違ったものが出てくるが, 背後にある統計的推測の問題として自然なゲームを設定することとミニマックス定理の成立がキーになる. このあたりの議論は, 数学的にはさほど難しくないも

のの統計的決定理論になじみがないと理解しづらい．（統計的決定理論とベイズとの関係は例えば Ferguson [13] を参照．）それらの解説は別の機会に行うこととして，本稿ではゲーム論的な解釈で得られる無情報事前分布 (MDP prior) の定義を紹介する．

Theorem 4.1
$\mathcal{H}$ を可分なヒルベルト空間とする．以下を満たす純粋状態モデル $\mathcal{M} = \{\rho_\theta : \theta \in \Theta \subseteq \mathbf{R}^k\}$ が与えられているとする．

(i) $\theta_1 \neq \theta_2 \Rightarrow \rho_{\theta_1} \neq \rho_{\theta_2}$.

(ii) Θ はコンパクト．

(iii) ρ_θ は θ について弱連続, i.e., 任意の有界作用素 X について, $\mathrm{Tr}\rho_\theta X$ が Θ 上, 連続関数．

この時, 以下を満たす事前分布 $\pi_* \in \mathcal{P}(\Theta)$ が存在する．

$$\|\rho_{\pi_*}\| = \inf_{\pi\in\mathcal{P}(\Theta)} \|\rho_\pi\|,$$

ここで, $\|X\|$ は X の作用素ノルムを表しており,

$$\rho_\pi := \int_\Theta \rho_\theta \pi(\mathrm{d}\theta)$$

と定義している．

Definition 4.2
上の設定で π_* を最大検出確率事前分布 (maximum detection probability prior) とよぶ．MDP prior と略記する．

コンパクト性を外すと存在しない例が容易に作られる．(詳しくは著者の原論文 [30] を参照．上の結果に関連するミニマックス定理の証明や MDP prior の具体例も原論文にある)．また, 一般に MDP prior も reference prior も量子系の場合には一意的には定まらない例が普通に出てくる．(単位作用素 I が異なる基底の混合でかけることを用いれば作れる.)

5 一般化エントロピーとの関係

前節でみたように MDP prior は $\|\rho_\pi\|$ を最小化する π として定義された．一方で, 純粋状態モデルの場合, Bernardo 流のやり方を形式的に拡張すると $S(\rho_\pi)$ の最

大化が出てくる．したがって，この点だけに注目すると，密度作用素 ρ の作用素ノルム $\|\rho\|$ とエントロピー $S(\rho)$ との比較は興味深い．例えば，$\dim\mathcal{H}=2$ では二つの方法は常に一致し，エントロピーが最大値 $S(\rho_\pi)=\log d,\ d=\dim\mathcal{H}$ をとるような場合にも両者は一致する [30].

実は $\|\rho\|$ は quantum Rényi entropy の極限と対応していることがわかる．まず，そのことを確認しておこう．以下，ρ の第一固有値（固有値で最大のもの）が縮退していないことを仮定する．

まず，古典情報理論におけるエントロピーの一般化として，Rényi entropy が古くから知られている [26]．これを密度作用素に拡張したものは quantum Rényi entropy とよばれ，以下で与えられる．

$$S_q^R(\rho) := \frac{\log \mathrm{Tr}\rho^q}{1-q}.$$

quantum Rényi entropy は量子情報理論でエンタングルメントを特徴づける量として時々使われる．また，近年は Tsallis entropy

$$S_q(\rho) := -\frac{1-\mathrm{Tr}\rho^q}{1-q}$$

も提案されて統計力学や情報幾何での研究が多い．([31]; ここでは密度作用素の場合で記載した.) 両者はともに $\mathrm{Tr}\rho^q$ を含み，$q\to 1$ とすると通常のエントロピーに帰着するため一般化エントロピーと考えられる．一方で，$q\to\infty$ としたときの極限に注目すると，Rényi entropy は有限の値に収束し

$$T(\rho) := \lim_{q\to\infty} S_q^R(\rho) = -\log\|\rho\|$$

が得られる．(Tsallis entropy は 0 になってしまう.)

以下，エントロピー $S(\rho)$ と $T(\rho)$ の対比を中心に見ていく．まず，有界自己共役作用素の最大固有値は作用素ノルムに一致するため

$$e^{-T(\rho)} = \|\rho\| = \max_{\varphi\in\mathcal{H}:\ \|\varphi\|=1} \langle\varphi|\rho|\varphi\rangle$$

と書けることに注意する．$T(\rho)$ は Rényi entropy の極限でかけるため，Rényi entropy で成立する性質の多くはそのまま受け継がれる．たとえば $T(\rho)$ は次の性質をもつことが容易に示される．

$$\begin{gathered} T(\rho\otimes\sigma) = T(\rho)+T(\sigma), \\ T(\lambda\rho+(1-\lambda)\sigma) \le \lambda T(\rho)+(1-\lambda)T(\sigma),\ 0\le\forall\lambda\le 1 \end{gathered}$$

以上から MDP prior は $T(\rho_\pi)$ の最大化から得られることがわかった. $S(\rho_\pi)$ の最大化を用いる場合とどのような条件下で一致するのだろうか. これに関しては, 現在, 検討中であるが, majorization を用いて両者が一致する十分条件を求めることができる. Majorization の定義, 及び, 以下の命題の証明は例えば Bhatia [6] を参照.

Proposition 5.1
各成分が非負実数値からなる二つのベクトルを $\vec{x} = (x_1, \dots, x_d)^\top$, $\vec{y} = (y_1, \dots, y_d)^\top$ とおく ($\top$ は転置を表す). この時, 以下は同値.

(i) $\vec{x} \preceq \vec{y}$

(ii) 二重確率行列 A が存在して $\vec{x} = A\vec{y}$,

(iii) ベクトルは成分を入れ替えたベクトルの凸結合でかける, つまり $x_j = \sum_\sigma p_\sigma y_{\sigma(j)}$.
(σ は置換群の元, p_σ は置換群の上の確率分布.)

Proposition 5.2
上の同値な条件を満たす時, 古典エントロピーについて $S(\vec{x}/\sum_j x_j) \geq S(\vec{y}/\sum_j y_j)$ が成立.

Definition 5.3
$d = \dim \mathcal{H} < \infty$ の時, $\rho, \sigma \in \mathcal{S}(\mathcal{H})$ について ρ の固有値ベクトルを $\vec{\lambda_\rho}$, σ の固有値ベクトルを $\vec{\lambda_\sigma}$ とあらわす. $\vec{\lambda_\rho} \preceq \vec{\lambda_\sigma}$ の時, $\rho \preceq \sigma$ と定義する.

以上を用いればただちに以下がわかる.

Lemma 5.4
(純粋状態族とは限らない) 一般の量子統計モデル $\mathcal{M} = \{\rho_\theta : \theta \in \Theta \subseteq \mathbf{R}^k\} \subset \mathcal{S}(\mathcal{H})$ について, 弱連続性を仮定 (ここでは Θ は一般の Borel 集合でよい.) この時, 以下が成立. $\forall \pi, \xi \in \mathcal{P}(\Theta)$,

$$\rho_\pi \preceq \rho_\xi \Rightarrow T(\rho_\pi) \geq T(\rho_\xi),\ S(\rho_\pi) \geq S(\rho_\xi).$$

以上を用いて本稿の主結果を得る.

Theorem 5.5
純粋状態族について, 以下の意味で一様に majorize される量子状態 ρ_π が作れたと

する.

$$\rho_\pi \preceq \rho_\xi,\ \forall \xi \in \mathcal{P}(\Theta).$$

この時, π は MDP prior かつ reference prior となる.

以上から互いに majorize されない状態が複数出てくる, つまり, 順序関係で極小な ρ_π が複数出てくる場合に MDP prior と reference prior で違う結果が得られると期待される. このようなケースについての詳細な解析と物理的な意味の考察や, 統計的推測との関連については, 別の機会に述べる予定である.

Acknowledgements

The author was supported by Kakenhi for Young Researchers (B) (No. 24700273).

REFERENCES

[1] J. Aitchison: Goodness of prediction fit. *Biometrika*, **62** (1975), 547–554.

[2] S. Amari and H. Nagaoka: *Methods of Information Geometry.* AMS, Oxford, 2000.

[3] L. M. Artiles, R. D. Gill, and M. I. Guţă: An invitation to quantum tomography. *J. R. Statist. Soc. B*, **vol.67** (2005), pp.109–134.

[4] J. M. Bernardo: Reference posterior distributions for Bayesian inference. *J. R. Statist. Soc. B*, **41** (1979), pp. 113–147.

[5] J. M. Bernardo: Reference analysis, In: Dey, K. K. ,Rao, C. R. (Eds.), *Handbook of Statistics*, vol.25. Elsevier, Amsterdam, 17–90, 2005.

[6] R. Bhatia: Matrix analysis. Springer, New York, 1997.

[7] C. M. Bishop: Pattern Recognition and Machine Learning. Springer-Verlag, New York, 2006.

[8] R. Blume-Kohout: Optimal, reliable estimation of quantum states. *New J. Phys.*, **12** (2010), 043034.

[9] V. Bužek, R. Derka, G. Adam and P. L. Knight: Reconstruction of quantum states of spin systems: from quantum Bayesian inference to quantum tomography. *Ann. Phys. (N.Y.)*, **266** (1998), 454–496.

[10] B. S. Clarke and A. R. Barron: Jeffreys' prior is asymptotically least favorable under entropy risk. *J. Statist. Planning and Inference*, **41** (1994), pp.36–60.

[11] T. M. Cover and J. A. Thomas: *Elements of Information Theory*. 2nd ed. Wiley-Interscience, New York, 2005.

[12] L. Davisson and A. Leon-Garcia: A source matching approach to finding minimax codes. *IEEE Trans. Inform. Theory*, **26** (1980), pp.166–174.

[13] T. Ferguson: *Mathematical Statistics: A Decision Theoretic Approach*. Academic Press, New York, 1967.

[14] P. D. Grünwald: *The Minimum Description Length principle*. MIT Press, Massachusetts, 2007.

[15] D. Haussler. A General Minimax Result for Relative Entropy. *IEEE Trans. Inform. Theory*, **43** (1997), pp.1276–1280.

[16] C. W. Helstrom: *Quantum Detection Theory*. Academic Press, New York, 1976.

[17] A. S. Holevo: Information-theoretic aspects of quantum measurement. *Problems of Inform. Transm.*, **9** (1973), pp.177–183.

[18] A. S. Holevo: *Probabilistic and Statistical Aspects of Quantum Theory*. North-Holland, Amsterdam, 1982.

[19] E. T. Jaynes: Information theory and statistical mechanics. I. *Phys. Rev.*, **106** (1957), pp.620–630.

[20] E. T. Jaynes: Information theory and statistical mechanics. II. *Phys. Rev.*, **108** (1957), pp.171–190.

[21] H. Jeffreys: Theory of Probability. Oxford University Press, Oxford, 1961.

[22] R. E. Kass and L. Wasserman: The selection of prior distributions by formal rules. *J. American Statistical Assoc.*, **vol. 91, no.435**, (1996), pp.1343–1370.

[23] M. A. Nielsen and I. L. Chuang: *Quantum Computation and Quantum Information.* Cambridge University Press, Cambridge, 2000.

[24] M. Ohya and D. Petz: *Quantum Entropy and Its Use.* Springer, Berlin, 2004.

[25] M. Paris and J. Řeháček: *Quantum State Estimation.* Springer, Berlin, 2004.

[26] A. Rényi: On measures of entropy and information. In *Proceedings of the 4th Berkeley Symposium on Mathematical Statistics and Probability*, Vol.1 (1961), University of California Press, pp.547–561.

[27] C. P. Robert: *The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation.* Springer, New York, 2001.

[28] 田中 冬彦: 統計学の新たなフロンティアとしての量子統計. 京都大学数理解析研究所共同研究集会 "量子論における統計的推測の理論と応用", 数理解析研究所講究録, 2013 年 (to appear).

[29] F. Tanaka and F. Komaki: Bayesian predictive density operators for exchangeable quantum-statistical models. *Phys. Rev. A*, **71** (2005), 052323.

[30] F. Tanaka: Noninformative prior in the quantum statistical model of pure states. *Phys. Rev. A*, **85** (2012), 062305.

[31] C. Tsallis: *Introduction to Nonextensive Statistical Mechanics*, Springer, New York, 2010.

[32] J. O. S. Yin and S. J. van Enk: Information criteria for efficient quantum state estimation. *Phys. Rev. A*, **83** (2011), 062110.